



Berner Fachhochschule  
Haute école spécialisée bernoise  
Bern University of Applied Sciences

# Refresher Workshop: Quantitative Methoden

# Über mich

## INTERNATIONALES MARKETING



## KREATIV



## DIGITAL

**Mit kreativen Schreibmethoden  
kompetent zur Abschlussarbeit**



Ein Kurs mit  
Prof. Dr. Christina Sichtmann



**Prof. Dr.  
Christina Sichtmann**

**Du schaffst das!**

Erfolgreich deine  
Abschlussarbeit schreiben

Der neue Podcast!  
Jetzt kostenlos abonnieren auf  
iTunes, Spotify oder  
[www.sichtmann.de](http://www.sichtmann.de)

# Podcast



**Prof. Dr.  
Christina Sichtmann**

**Du schaffst das!**  
Erfolgreich deine  
Abschlussarbeit schreiben

**Der neue Podcast!**  
Jetzt kostenlos abonnieren auf  
iTunes, Spotify oder  
[www.sichtmann.de](http://www.sichtmann.de)



# Refresher Workshop: Umfragen

Prof. Dr. Christina Sichtmann

► Departement W | Institute Applied Data Science & Finance

# Grundlagen zu Umfragen (Surveys)

- ▶ Die Umfrage (Befragung) ist neben der Beobachtung und dem Experiment eines der wichtigsten Instrumente der Sozialwissenschaften.
- ▶ Sie ist eine wissenschaftliche Methode zur Erhebung von Informationen. Welche Form der Befragung zur Anwendung kommt, ist abhängig vom Untersuchungsgegenstand und davon, welche Art von Informationen benötigt werden.

# Typischer (quantitativer) Survey-Forschungsprozess

## 1) Auswahl der Variablen

- Welche Variablen müssen gemessen werden, um Zusammenhänge zu analysieren?

## 2) Auswahl der Methodik

- Online, Persönlich, Schriftlich?

## 3) Gestaltung des Fragebogens

- Welche Fragen messen die Variablen von Interesse?
- Reihenfolge?

## 4) Datenerhebung

- Grösse der Stichprobe?
- Ideale Stichprobenstruktur?
- Wie erreicht man eine hohe Rücklaufquote?

## 5) Evaluation der Messung

- Messen die Survey-Fragen, was sie messen sollen?

# Questionnaire Design



Survey-  
Inhalt

- ▶ Konzentration auf Wichtiges (Dilemma: Vergessene «Z» können in aller Regel nicht nacherhoben werden)
- ▶ Daumenregel maximale Länge
  - ▶ Sollte idealerweise 15 Minuten nicht überschreiten (ideal: 5-10 Minuten)
- ▶ Vorgehensweise:
  - ▶ Basis: Literaturrecherche (Skalen)
  - ▶ Sinnvoll: Priorisierung der Wichtigkeit der Fragen



# Questionnaire Design

Survey-  
Inhalt

Frage-  
format

Bitte beschreiben Sie eine Situation, in der Sie sich in dieser Woche glücklich gefühlt haben.

Wie fühlen Sie sich gerade?

Überhaupt  
nicht glücklich

Sehr  
glücklich

1 2 3 4 5 6 7

☐ ☐ ☐ ☐ ☐ ☐ ☐

## Offene Fragen

Teilnehmer/innen können so antworten, wie sie wollen

Vorteile:

- Ermöglicht ungewöhnliche Antworten
- Geringes Vorwissen des/der Forschenden nötig

Nachteile:

- Schwierig quantitativ zu analysieren
- Erfordert mehr Mühe des Befragten, besonders bei schriftlichen Surveys

## Geschlossene Fragen

Teilnehmer/innen wählen aus vorgegebenen Antwortkategorien

Vorteile:

- Einfach quantifizierbar (kodierbar) und vergleichbar
- Zeiteffizient

Nachteile

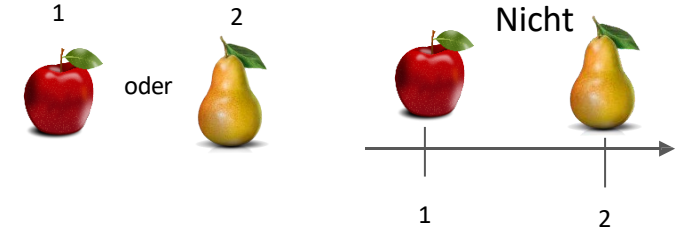
- Verlust der Spontanität
- Vollständigkeit der Antwortkategorien ggf. schwierig zu erreichen



# Skalentypen bei geschlossenen Fragen

## Nominal

- nicht – kontinuierlich
- Objekte werden lediglich klassifiziert



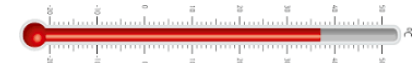
## Ordinal

- Objekte werden in Relation zueinander gesetzt
- Die Abstände zwischen den Objekten werden nicht berücksichtigt



## Intervall quasi-metrisch

- Die Abstände der Objekte können miteinander verglichen werden
- Es besteht kein natürlicher Nullpunkt



## Metrisch auch Ratio

- Es gibt einen natürlichen Nullpunkt
- Verhältnisse zwischen den Skalenwerten können berechnet werden



# Skalentypen bei geschlossenen Fragen

Skala	Beschreibung	Beispiel	Auswertung
<b>Nominal</b>	Identifikation und / oder Klassifikation der Objekte	Marke, Geschlecht, Einkaufsort	Prozent, Chi-Quadrat
<b>Ordinal</b>	Bilden von Rangordnungen ohne vergleichbare Abstände zwischen den Objekten	Marktposition, Schulnoten	Prozent
<b>Intervall</b>	Vergleich der Abstände zwischen den Objekten; willkürlicher Nullpunkt	Likert-Skala, Kaufabsicht	Mittelwert, t-Test, ANOVA, Regression
<b>Metrisch</b>	Natürlicher Nullpunkt; ermöglicht einen Vergleich der Abstände und deren Verhältnisse	Alter, Umsatz, Einkommen	Mittelwert, t-Test, ANOVA, Regression

# Geschlossene Fragen – Alternativfrage

**Alternativfragen** verfügen grundsätzlich nur über zwei Antwortkategorien, z. B. „ja / nein“ oder „stimme zu / stimme nicht zu“.

Unter Umständen kann es sinnvoll sein eine neutrale Alternative wie z.B. „weiss nicht“, „weder noch“ bzw. „sowohl als auch“ anzubieten.

Bitte geben Sie Ihr Geschlecht an:

- ☐ weiblich
- ☐ männlich
- ☐ divers

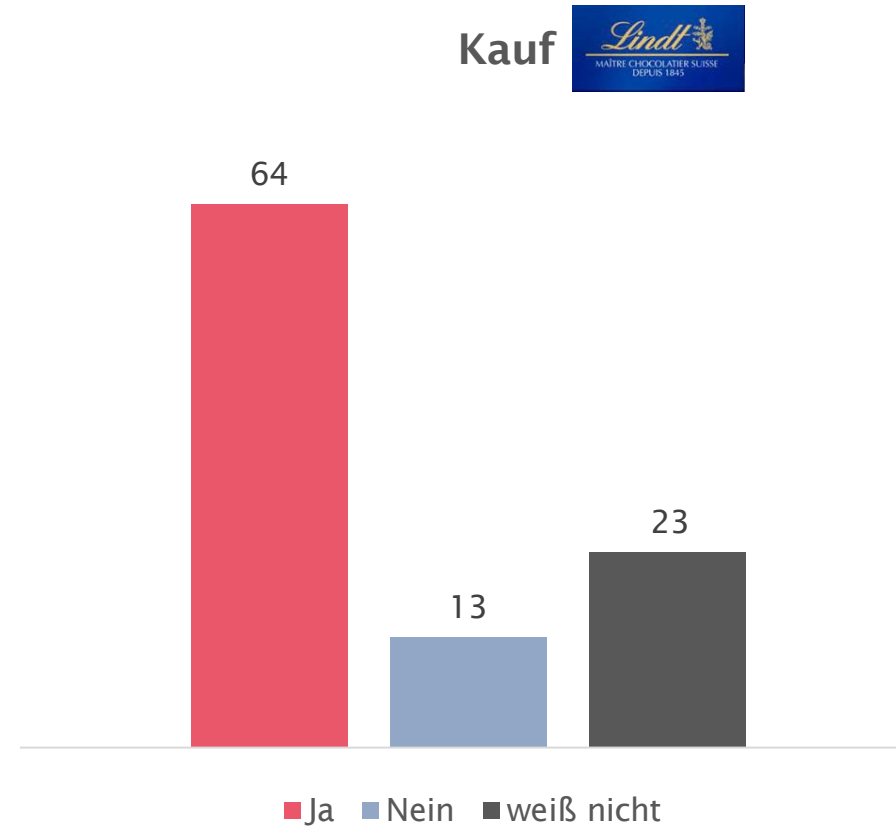
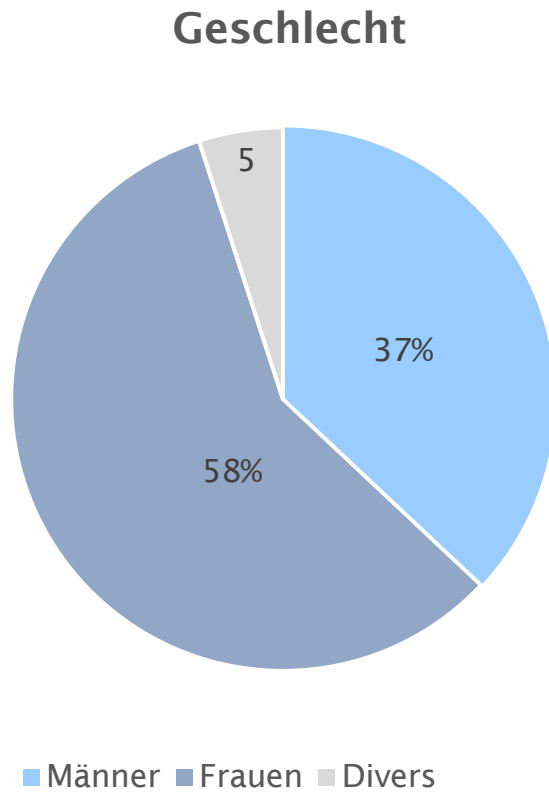
Haben Sie in den vergangen sechs Wochen Schokolade der Marke



gekauft?

- ☐ ja
- ☐ nein
- ☐ weiss nicht

# Auswertung Alternativfragen



# Geschlossene Fragen – Multiple Choice Frage

**Multiple Choice Fragen** zeichnen sich dadurch aus, dass sie mehrere alternative Antwortkategorien zulassen. Die Anzahl der möglichen Nennungen kann dabei begrenzt oder unbegrenzt sein.

Von welcher der folgenden Marken haben Sie in den letzten sechs Monaten Schokolade gekauft?

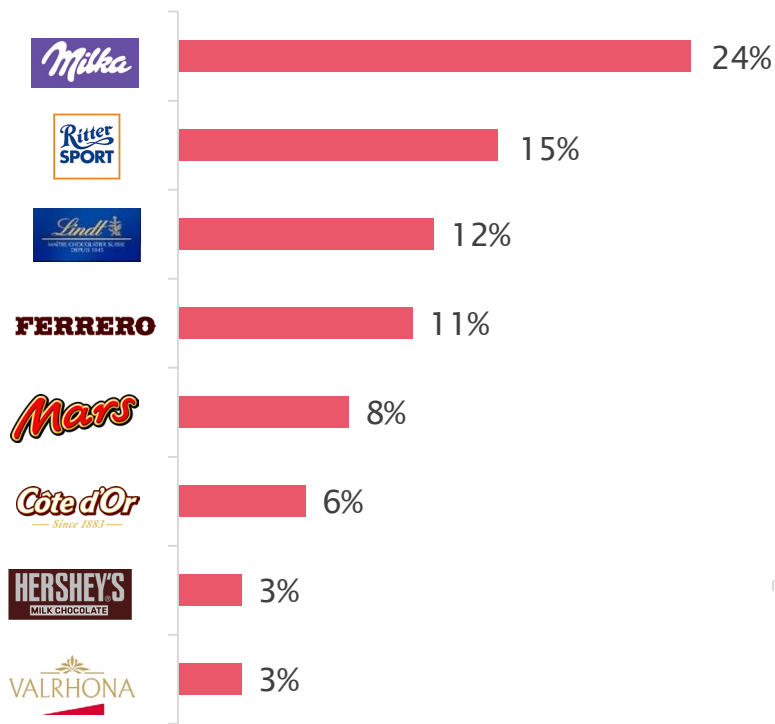
☐☐☐☐☐☐☐☐

Welche der folgenden Kriterien spielen beim Kauf von Tomaten für Sie eine Rolle?

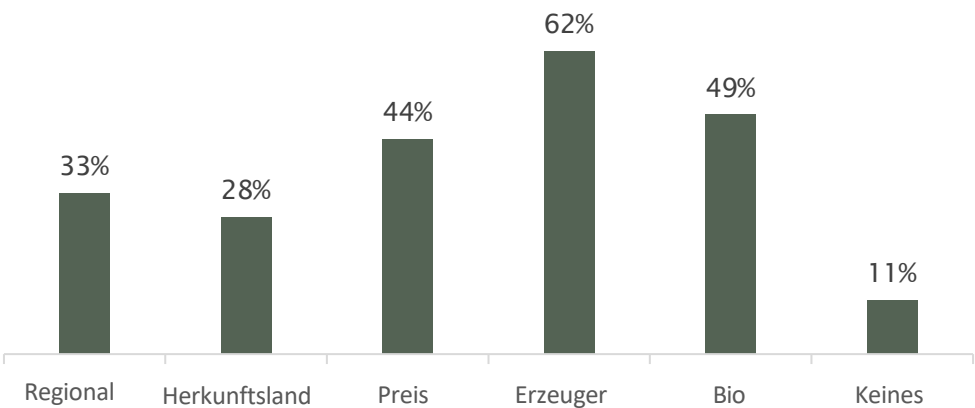
- ☐ Regionaler Anbau
- ☐ Herkunftsland
- ☐ Preis
- ☐ Erzeuger
- ☐ Biologischer Anbau
- ☐ Keines dieser Kriterien

# Auswertung von Multiple Choice Frage

Schokolade in den letzten 6 Monaten



Kaufkriterien Tomaten



# Likert Skalen

Befragte geben an, inwiefern Sie einer Aussage zustimmen, oder ein Merkmal für zutreffend halten.  
Gängige Skalen sind 5Pkt. oder 7Pkt.-Skalen

Im Folgenden sehen Sie eine Auswahl an Aussagen, die andere über das Unternehmen  
Bitte geben Sie an, inwieweit Sie diesen Aussagen zustimmen.

MONSANTO  getroffen haben.

	Stimme gar nicht zu						Stimme voll und ganz zu	
Monsanto würde zugunsten der Umwelt auch Umsatzeinbussen in Kauf nehmen	[1]	<del>[2]</del>	[3]	[4]	[5]	[6]	[7]	
Monsanto legt großen Wert auf die Unbedenklichkeit von Produkten	[1]	[2]	<del>[3]</del>	[4]	[5]	[6]	[7]	
Monsanto ist ein Vorreiter in puncto Corporate Social Responsibility	[1]	<del>[2]</del>	[3]	[4]	[5]	[6]	[7]	
Wenn ich an Monsanto denke fällt mir ausschließlich Positives ein	<del>[1]</del>	[2]	[3]	[4]	[5]	[6]	[7]	
Wären alle Unternehmen wie Monsanto, wäre die Welt ein besserer Ort	[1]	[2]	<del>[3]</del>	[4]	[5]	[6]	[7]	



# Questionnaire Design



Anzahl  
Items

- ▶ Wie viele Fragen / Items sollten eine Variable messen?
  - ▶ Single-Item-Messung (Berkvist und Rossiter 2007)
  - ▶ Multi-Item-Messung (Diamantopoulos et al. 2012)
- ▶ Manifeste Variablen (Alter, Einkommen, Geschlecht, etc.) einfach durch eine Frage messbar (Z.B.: Wie alt sind Sie?)
- ▶ Für latente Variablen/Konstrukte (z.B. Einstellungen, Persönlichkeitseigenschaften, Intentionen) werden in der Regel Multi-Item-Skalen verwendet
  - ▶ Idee: Ein Set mehrerer Items misst dasselbe Konstrukt reliabler als die individuellen Items
  - ▶ Geteilte Varianz der Items «reflektiert» das zugrunde liegende Konstrukt
  - ▶ Abnehmender Grenznutzen zusätzlicher Items, ideal sind 4-5

# Questionnaire Design

## Item-Inhalt

### ► Beispiele:

#### Construct

##### Consumer innovativeness<sup>a</sup>

In general, I am among the first in my circle of friends to buy a consumer electronic product when it appears.

Compared to my friends, I often shop for consumer electronic products.

If I heard that a new consumer electronic product was available through a local department store, I would be interested enough to buy it.

In general, I am the first in my circle of friends to know the names of the latest consumer electronic product.

I know about consumer electronic products before other people do.

##### Price fairness<sup>b</sup>

*I personally feel that the advertised selling price for the \_\_\_ is...*

Unfair–fair

Unreasonable–reasonable

Unacceptable–acceptable

Unjustified–justified

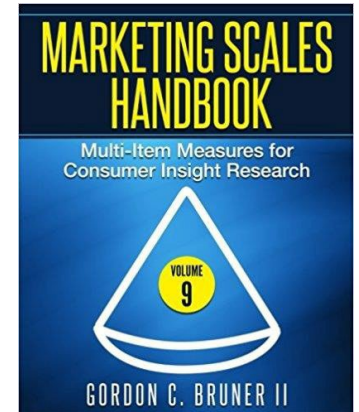
<sup>a</sup>Adopted from Goldsmith and Hofacker (1991); 1 = “strongly disagree”, 7 = “strongly agree”.

<sup>b</sup>Adopted from Bolton et al. (2010)/Vaidyanathan and Aggarwal (2003); 1 = “unfair/unreasonable/unacceptable/unjustified”, 7 = “fair/reasonable/acceptable/justified”.

# Questionnaire Design

## Item-Inhalt

- ▶ Entwicklung eines Messinstruments für latente Konstrukte extrem aufwändig
- ▶ Deshalb: Verwendung etablierter Skalen
- ▶ Bücher mit etablierten Skalen sind ebenfalls verfügbar
- ▶ Gegebenenfalls Anpassungen:
  - ▶ Auf forschungsspezifischen Kontext
  - ▶ Möglichst inhaltsgenaue Übersetzung auf Deutsch
- ▶ Mitunter existieren verschiedene Messinstrumente für dasselbe Konstrukt. Wählen Sie die Skala, die inhaltlich am ehesten Ihre Definition widerspiegelt



# Questionnaire Design



## Item-Inhalt

### ► Sonstige allgemeine Tipps zur Formulierung von Fragen

Einfach!	Klar!	Neutral!
<ul style="list-style-type: none"><li>• Kurze Fragen (max. 15 Wörter)</li><li>• Keine Relativsätze</li><li>• Vermeidung von Negationen</li><li>• Vermeidung von wissenschaftlichem Jargon (Variablennamen, etc.)</li></ul>	<ul style="list-style-type: none"><li>• Vermeidung der Begriffe «immer», «alles»</li><li>• Vermeidung vager Begriffe wie «oft», «viele»</li><li>• Definitionen von Fragen trennen</li><li>• Angabe eines Zeithorizonts, auf den sich die Frage bezieht</li><li>• Pro Frage nur auf einen Aspekt beschränken</li></ul>	<ul style="list-style-type: none"><li>• Keine Suggestivfragen</li><li>• Keine Antizipation von Antworten («Sind Sie nicht auch der Meinung, dass...»)</li><li>• Vermeidung aufgeladener Begriffe</li></ul>

# Questionnaire Design



- ▶ Vom Allgemeinen zum Speziellen
- ▶ Demografika am Ende, ausser sie werden benötigt, um die richtige Stichprobe zu finden (Screening-Kriterien)
- ▶ Einfache, interessante Fragen zu Beginn
- ▶ Wichtige Fragen zu Beginn (die für Hypothesentests benötigt werden)
- ▶ Handlungsbezogene Fragen (Auswahl, Entscheidungen, Intentionen) vor Einstellungen, Wahrnehmungen
- ▶ Heikle Fragen später im Fragebogen
- ▶ Bei Verzweigungen: Klare Anweisungen (online: entsprechende Programmierung)
- ▶ Aufgepasst: Probanden formen oft Einstellungen im Verlauf des Fragebogens -> frühe Fragen aktivieren Informationen, die dann die Beantwortung späterer Fragen beeinflussen

# Questionnaire Design



- ▶ Questionnaire should look as sharp as your résumé (Churchill and Iacobucci 2005, p. 252)
  - ▶ Signalisiert Interesse für das Thema
  - ▶ Signalisiert Vertrauenswürdigkeit des Durchführenden
  - ▶ Verringert die Abbruchquote
- ▶ Elemente des Layouts:
  - ▶ Klar, einfach, übersichtlich
  - ▶ Klare Identifizierung der Forschenden
  - ▶ Lesbare Schrift, ausreichende Schriftgrösse
  - ▶ Klare Überleitungen, wichtige Erklärungen und verständliche Instruktionen
- ▶ Ansprechende Einladung formulieren und sich für die Teilnahme bedanken!

# Questionnaire Design



- ▶ Daumenregel: Probanden verstehen alles falsch, was falsch zu verstehen ist
- ▶ Test des Fragebogens mit wenigen (<5) Probanden
- ▶ Ziele:
  - ▶ Identifikation von Ambiguitäten und Unklarheiten
  - ▶ Identifikation von fehlendem Vorwissen, um Fragen beantworten zu können
  - ▶ Überprüfung der Vollständigkeit von Antwortkategorien
  - ▶ Messung der Zeit der Bearbeitung





# Refresher Workshop: Experimente

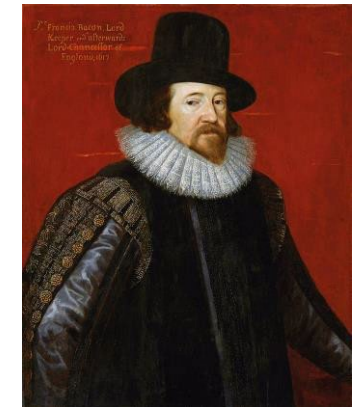
Prof. Dr. Christina Sichtmann

► Departement W | Institute Applied Data Science & Finance

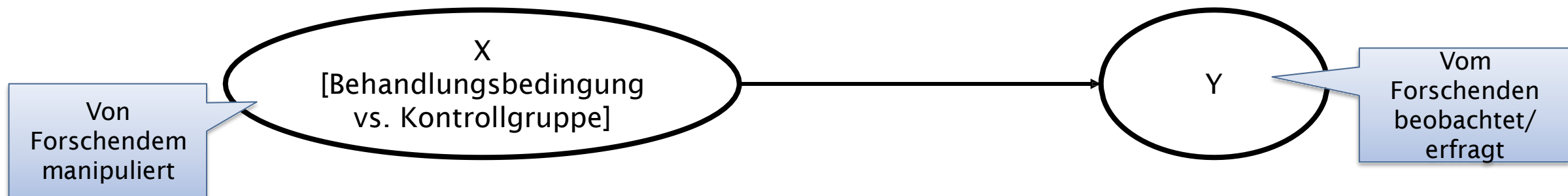
# Alternative zu Umfragen: Experimentelle Manipulation

„An experiment is taken to mean a scientific investigation in which an investigator manipulates (...) one or more independent variables and observes the dependent variable or variables for variation concomitant to the manipulation of the independent variables.“ (Churchill/Iacobucci 2005, S. 128)

- ▶ Aktive Manipulation der X-Variable(n) ist zentrales Unterscheidungsmerkmal zu anderen Erhebungsdesigns
- ▶ Y-Variable: Weiterhin Beobachtung/Befragung
- ▶ Zentrale Innovation der “wissenschaftlichen Revolution” im 17. Jahrhundert
- ▶ Francis Bacon: “not only must we observe nature in the raw, but we must also ,twist the lion’s tale‘, that is, manipulate our world in order to learn its secrets“ (as cited by Hacking 1983, p. 149)



Francis Bacon  
(1561-1626)



# Arten von Experimenten und Experimentaldesigns

## Randomisierung

«echtes» Experiment

Quasi-Experiment

vollständig

keine

## Umgebung

Laborexperiment

Szenario-  
Experiment

Physische  
Nachstellung

Feldexperiment

hypothetisch

realistisch

## Forschungsdesign: Unterschiede zwischen oder Innerhalb der Gruppen

«Between-Subjects»-Design

«Within-Subjects»-Design

Unterschiedliche Teilnehmer  
in den Gruppen

Gleiche Teilnehmer in den  
Gruppen

## Forschungsdesign: Anzahl der UV und ihrer Ausprägungen

Wie viele UV werden  
in wie vielen Stufen  
manipuliert?

# Echtes Experiment vs. Quasi-Experiment

## Echtes Experiment:

- ▶ vollständige Randomisierung → zufällige Zuordnung von Teilnehmern zu Experimentalgruppen



Zuteilung nach Zufallsprinzip

Gruppe 1



Gruppe 2



## Quasi-Experiment:

- ▶ Experiment mit manipulierter UV, aber keine randomisierte Zuteilung zu den Gruppen



Teilnehmende ordnen sich selbst einer Gruppe zu oder  
Leiter/in des Experiments ordnet gezielt  
Teilnehmende einer Gruppe zu

Gruppe 1

freiwillige  
Teilnahme am  
Bonusprogramm



Gruppe 2

keine Teilnahme  
am  
Bonusprogramm

# Labor- versus Feldexperimente



## Experimente

### Laborexperiment

#### Merkmale

- Umfeld speziell für Durchführungen von Laborexperimenten
- Teilnehmer/innen sind sich der Teilnahme bewusst (aber in Unkenntnis der Hypothesen)

#### Vorteile

- Relativ geringer Ressourcenaufwand
- Hohe Kontrolle über Manipulation der UV
- Relativ hohe Kontrolle über potenziell konfundierende Variablen (= **hohe interne Validität**)
- Relativ hohe Replizierbarkeit

#### Nachteile

- Künstliche Entscheidungssituation (= **niedrige externe Validität**)
- Teilnehmer/innen verhalten sich anders, weil sie wissen, dass sie an einem Experiment teilnehmen (= **niedrige externe Validität**)

### Feldexperiment

#### Merkmale

- Realistisches Umfeld (z.B. Migros, etc.)
- Teilnehmer/innen sind sich der Teilnahme meist nicht bewusst

#### Vorteile

- Hohe Authentizität der Entscheidungssituation erhöht Generalisierbarkeit der Ergebnisse (= **hohe externe Validität**)
- Teilnehmer/innen zeigen ihr «echtes» Verhalten

#### Nachteile

- Hoher Ressourceneinsatz
- Ethisch problematisch
- Relativ geringe Replizierbarkeit
- Relativ geringe Kontrolle über konfundierende Variablen (= **niedrige interne Validität**)
- Manipulation der UV oft problematisch

# Externe Validität in Laborsettings

- ▶ Wählen Sie Settings, Stimuli etc., die für Ihre Zielgruppe möglichst realistisch ist!
  - ▶ Z.B. Nachstellen einer Online-Kaufumgebung
  - ▶ Tradeoff zwischen mehr Realismus und mehr Ablenkung

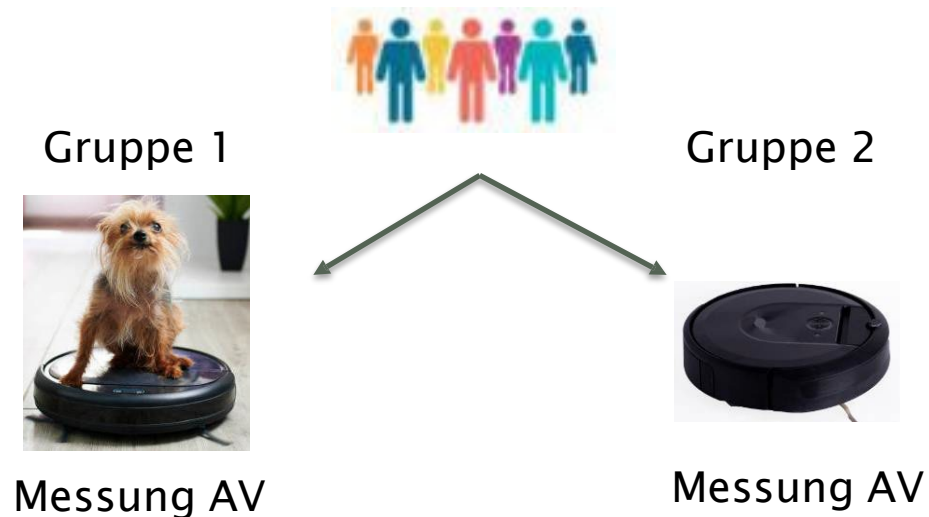


- ▶ Messen Sie (ggf. zusätzlich) möglichst tatsächliches Verhalten
  - ▶ Z.B. statt (nur) «Interesse» (auch zusätzlich) «Verweildauer auf Produktinformationen
  - ▶ Z.B. statt (nur) «negatives WoM» (auch zusätzlich) «tatsächliches Verfassen eines Kommentars»

# Between- vs. Within-Subject Designs

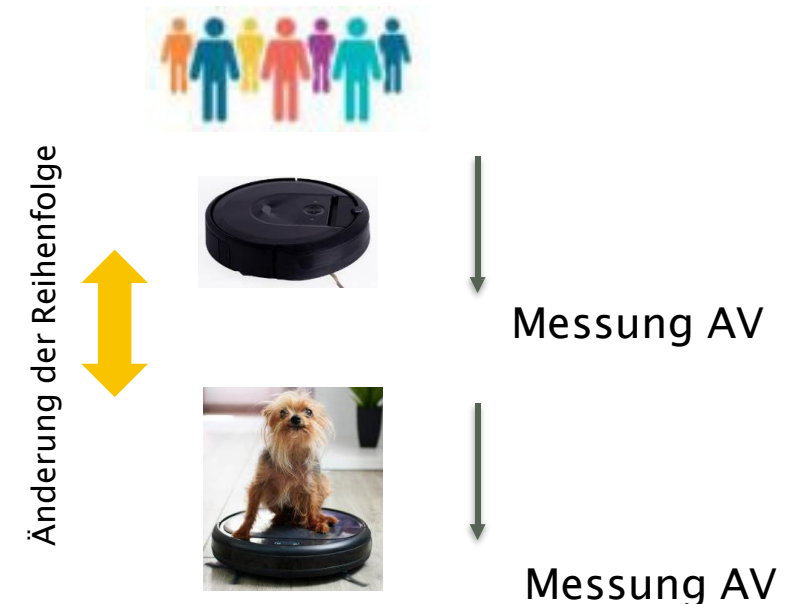
## Between-Subjects:

- ▶ Jeder Proband nimmt nur an einer Experimentalbedingung teil, woraus sich mehrere (min. 2) Gruppen ergeben, die miteinander verglichen werden
- ▶ Man ist hier an **Unterschieden zwischen verschiedenen Teilnehmern aus unterschiedlichen Gruppen** interessiert



## Within-Subjects:

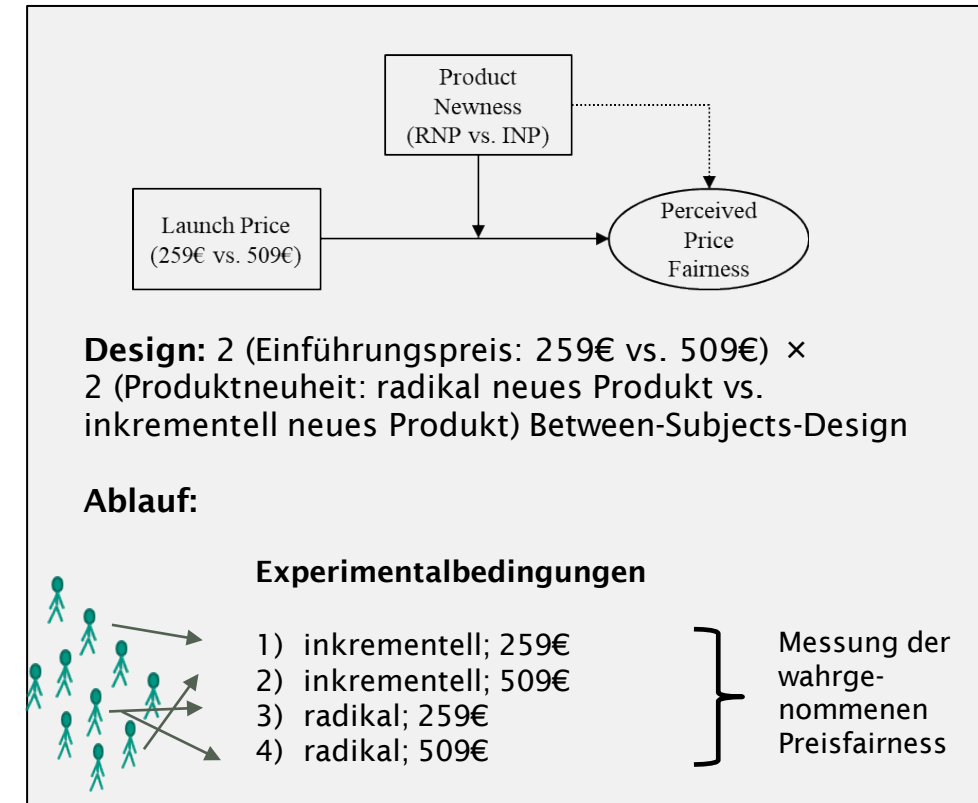
- ▶ Jeder Proband durchläuft alle Experimentalbedingungen nacheinander, dabei wird die AV nach jedem Durchlauf gemessen
- ▶ Es wird ein **Vergleich von Teilnehmern innerhalb einer Gruppe** angestellt









# Forschungsdesign: Manipulation mehrerer Variablen (I)

- ▶ Es ist möglich, mehr als eine UV zu manipulieren (z.B. zusätzlich ein Moderator)
  - ▶ In voll-faktoriellen Designs erhält man bei der (Varianz-) Analyse die direkten Effekte aller manipulierten UV auf die AV sowie alle möglichen Interaktionseffekte (ggf. höherer Ordnung)
  - ▶  $\# \text{ Experimentalbedingungen} = \# \text{ Bedingungen } UV_1 \times \# \text{ Bedingungen } UV_2 \dots \times \# \text{ Bedingungen } UV_N$
  - ▶ Beispiel: Hängt der Effekt des Einführungspreises auf die wahrgenommene Preisfairness von der Neuheit des Produkts ab? (Kuester et al., 2015)

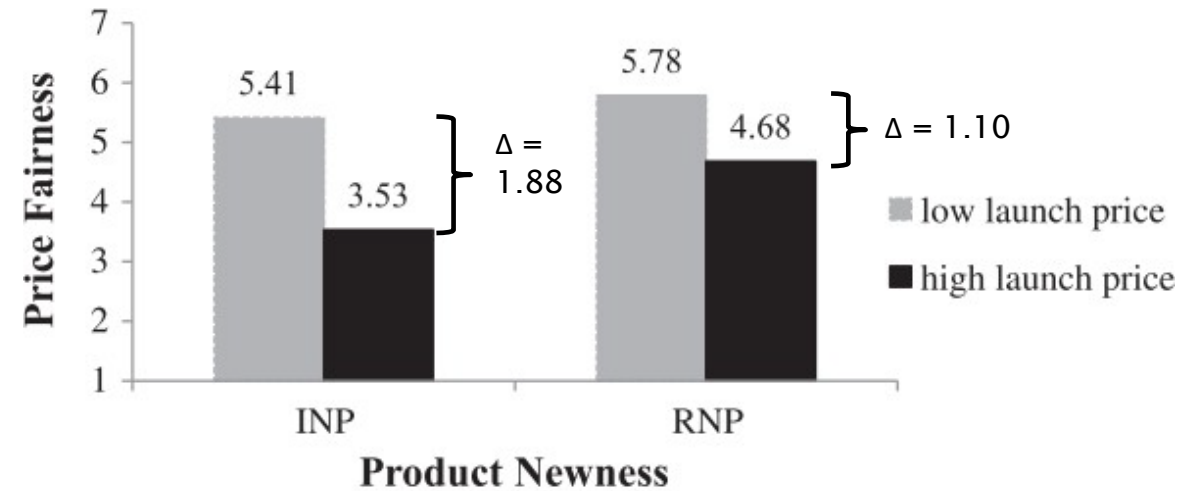


# Forschungsdesign: Beispiel Manipulation mehrerer Variablen

		Produktneuheit	
		INP	RNP
Einführungspreis	Niedrig	 259€ <ul style="list-style-type: none"><li>• 4.8", 1280x720 touch screen display</li><li>• Quadcore 1.4GHz processor</li><li>• 8 MP camera</li></ul>	 259€ <ul style="list-style-type: none"><li>• 4.8", 1280x720 touch screen display</li><li>• bendable OLED display technology</li><li>• Quadcore 1.4GHz processor</li><li>• 8 MP camera</li></ul>
	Hoch	 509€ <ul style="list-style-type: none"><li>• 4.8", 1280x720 touch screen display</li><li>• Quadcore 1.4GHz processor</li><li>• 8 MP camera</li></ul>	 509€ <ul style="list-style-type: none"><li>• 4.8", 1280x720 touch screen display</li><li>• bendable OLED display technology</li><li>• Quadcore 1.4GHz processor</li><li>• 8 MP camera</li></ul>

→ Zentraler Schritt: Gestaltung der Szenarien/Experimentalbedingung unter Sicherstellung dass sie sich *ausschliesslich* hinsichtlich der Manipulation unterscheiden (nicht immer trivial)

# Forschungsdesign: Beispiel Manipulation mehrerer Variablen

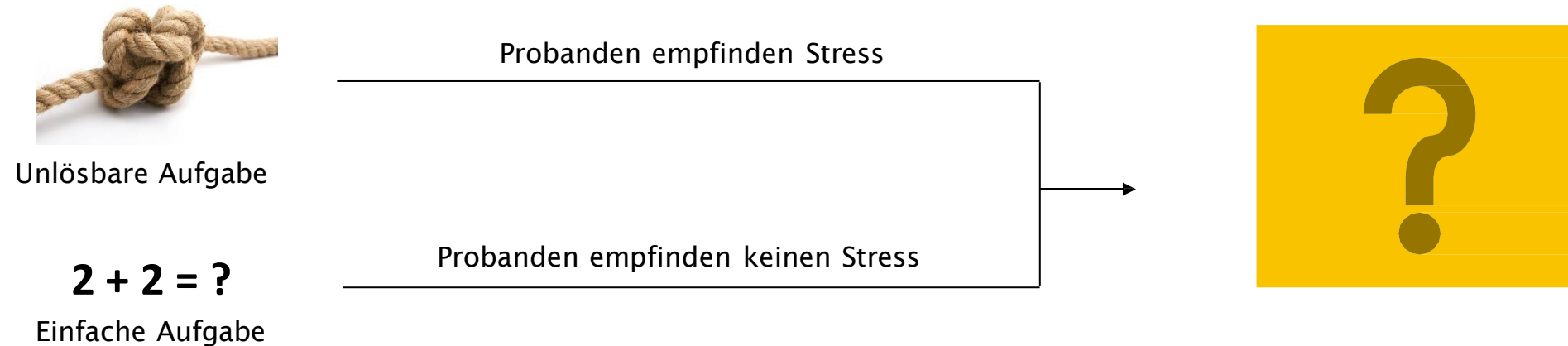


**Fig. 2.** Study 1: the launch price  $\times$  product newness interaction.

(Kuester et al. 2015)

# Wichtig bei Experimenten: Hat die Manipulation funktioniert?

**z. B. Manipulation von Frustration als unabhängige Variable mit den Stufen „nicht frustriert“ und „stark frustriert“**



Wie können wir nun aber sicher sein, dass unsere Manipulation funktioniert hat und die Teilnehmenden der beiden Gruppen tatsächlich spürbar unterschiedliche Frustrationslevel empfinden?

# Lösung: Manipulation Check

- ▶ Überprüft meist mit einer einfachen Frage die Effektivität der Manipulation.
- ▶ Inwiefern stimmen Sie der folgenden Aussage zu? (hier: sieben-Punkt Likert-Skala)
- ▶ *Ich fühle mich sehr gestresst (1) stimme überhaupt nicht zu ----- (7) stimme voll zu*
- ▶ Die Gruppenmittelwerte der Antworten auf diese Frage sollten sich zwischen den Gruppen signifikant unterscheiden

$\mu=6.1$



Unlösbare Aufgabe

$2 + 2 = ?$

$\mu=2.3$

Einfache Aufgabe

- ▶ Ggf. Test von Manipulationen im Rahmen von Pilotstudien (Pre-Test) nötig.
- Es empfiehlt sich für eigene Experimente auf bereits bewährte Manipulationen zu setzen und diese ggf. nur leicht für die eigene Fragestellung anzupassen.
- Ein Manipulationscheck ist allerdings auch dann nötig!



# Refresher Workshop: Datenanalyse

Prof. Dr. Christina Sichtmann




► Departement W | Institute Applied Data Science & Finance

# Agenda

- ▶ **Multi-Item Skalen**
- ▶ Festlegen der geeigneten statistischen Methode
  - ▶ Hypothesentests als Signifikanztests
  - ▶ t-Test
  - ▶ Einfaktorielle ANOVA (1 UV mit 2 oder mehr Stufen)



# Prüfung der internen Konsistenz einer Multi-Item Skala (I)

				...
Ich bin insgesamt zufrieden mit mir selber.	2	1	2	...
Manchmal habe ich das Gefühl, ich genüge überhaupt nicht.	1	4	2	...
Ich besitze eine Reihe guter Eigenschaften.	3	2	4	...
Ich kann Vieles genauso gut wie die meisten anderen Menschen auch.	3	4	4	...
Es gibt wenig, auf das ich stolz sein kann.	2	4	2	...
Ich fühle mich von Zeit zu Zeit nutzlos.	1	3	3	...
Ich halte mich für einen wertvollen Menschen.	3	3	1	...
Ich wünschte, ich hätte mehr Respekt vor mir selbst.	4	2	2	...
Alles in allem neige ich dazu mich für einen Versager zu halten.	3	2	3	...
Ich finde mich gut, so wie ich bin.	2	3	4	...

Konstrukt  
Selbstwertgefühl

4 = Ich stimme sehr zu  
3 = ich stimme zu  
2 = Ich stimme nicht zu  
1 = Ich stimme überhaupt nicht zu

Cronbachs Alpha ( $0 < \text{Alpha} < 1$ ) gibt das Ausmass an, indem die Items einer Skala miteinander in Beziehung stehen und somit geeignet sind ein und dasselbe Konstrukt abzubilden.

# Prüfung der internen Konsistenz einer Multi-Item-Skala (II)

- ▶ Schritt 1: Kodierung aller Items so, dass hohe Werte für das Item hohe Werte des zu messenden Konstrukts reflektieren (ggf. umkodieren)
- ▶ Schritt 2: SPSS: Analysieren → Metrisch → Reliabilitätsanalyse (unter „Statistiken: Haken bei Skale, wenn Item gelöscht)
- ▶ Schritt 3: Ist Alpha > 0.7?
  - ▶ Ja Berechnung des Durchschnitts für das zu messende Konstrukts als Mittelwert der jeweiligen Item-Werte
  - ▶ Nein Schrittweises Löschen der Items mit niedrigster Item-Total-Korrelation. Wiederholen bis Alpha > .7 oder nur noch 3 Items übrig. Dann Absprache mit Betreuenden. Reliabilitätsproblem! Ansprache in den Limitationen nötig.

Reliabilitätsanalyse: Statistik

Deskriptive Statistiken für

- ☐ Item
- ☐ Metrisch
- ☒ Skala, wenn Item gelöscht

Zwischen Items

- ☐ Korrelationen
- ☐ Kovarianzen

Auswertungen

- ☒ Mittelwerte
- ☒ Varianzen
- ☒ Kovarianzen
- ☒ Korrelationen

ANOVA-Tabelle

- ☒ Ohne
- ☐ E-Test
- ☐ Friedman-Chi-Quadrat
- ☐ Cochran-Chi-Quadrat

Bewerterübergreifende Übereinstimmung: Fleiss-Kappa

- ☐ Übereinstimmung bei einzelnen Kategorien anzeigen
- ☐ Zeichenfolgefälle ignorieren
- ☒ Kategoriebeschriftungen als Zeichenfolgen werden in Großbuchstaben angezeigt

Niveau für asymptotische Signifikanz (%): 95

Fehlend




- ☒ Benutzer- und systemdefiniert fehlende Werte ausschließen

Reliability Statistics				
Cronbach's Alpha	N of Items			
.896	4			

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
To what extent do you agree with the following questions about health? - I reflect about my health a lot.	18,41	22,264	,815	,848
To what extent do you agree with the following questions about health? - I'm very self-conscious about my health.	18,51	22,867	,790	,858
To what extent do you agree with the following questions about health? - I'm constantly examining my health.	18,69	21,988	,816	,848
To what extent do you agree with the following questions about health? - I'm alert to changes in my health.	18,07	27,082	,667	,902

# Berechnung des Skalenmittelwertes über alle Items hinweg für jeden Probanden

				...
Ich bin insgesamt zufrieden mit mir selber.	2	1	2	...
<b>Ich habe das Gefühl, ich genüge (rec.).</b>	4	1	3	...
Ich besitze eine Reihe guter Eigenschaften.	3	2	4	...
Ich kann Vieles genauso gut wie die meisten anderen Menschen auch.	3	4	4	...
Es gibt wenig, auf das ich stolz sein kann.	2	4	2	...
Ich fühle mich von Zeit zu Zeit nutzlos.	1	3	3	...
Ich halte mich für einen wertvollen Menschen.	3	3	1	...
Ich wünschte, ich hätte mehr Respekt vor mir selbst.	4	2	2	...
Alles in allem neige ich dazu mich für einen Versager zu halten.	3	2	3	...
Ich finde mich gut, so wie ich bin.	2	3	4	...
<b>Selbstwert_Mean</b>	<b>2,7</b>	<b>2,5</b>	<b>2,8</b>	<b>...</b>

Wurde vor der Bildung des Mittelwerts recodiert!!!

**Konstrukt Selbstwertgefühl**

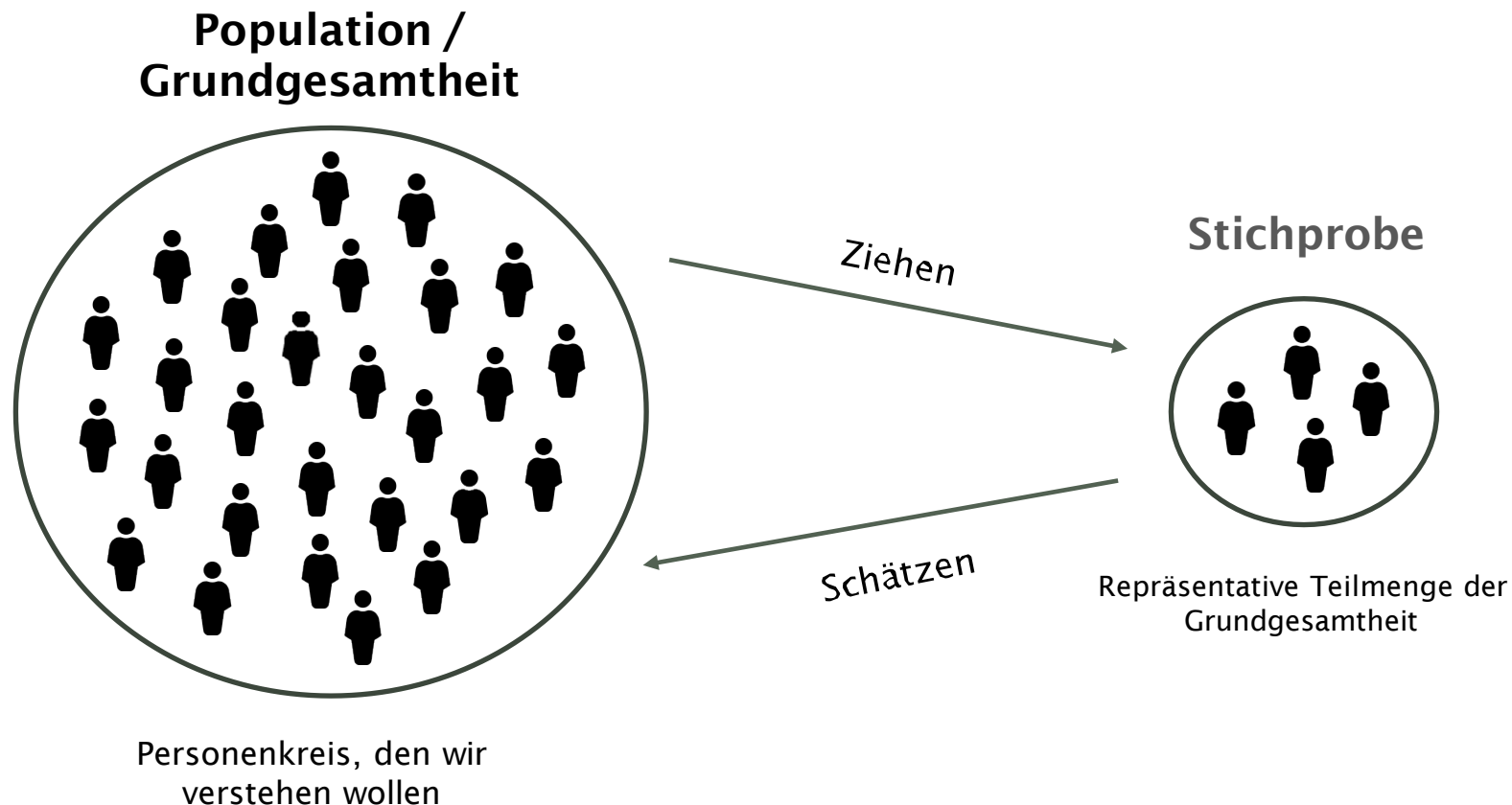
4 = Ich stimme sehr zu  
 3 = ich stimme zu  
 2 = Ich stimme nicht zu  
 1 = Ich stimme überhaupt nicht zu

# Agenda

- ▶ Multi-Item Skalen
- ▶ Festlegen der geeigneten statistischen Methode
  - ▶ **Hypothesentests als Signifikanztests**
  - ▶ t-Test
  - ▶ Einfaktorielle ANOVA (1 UV mit 2 oder mehr Stufen)

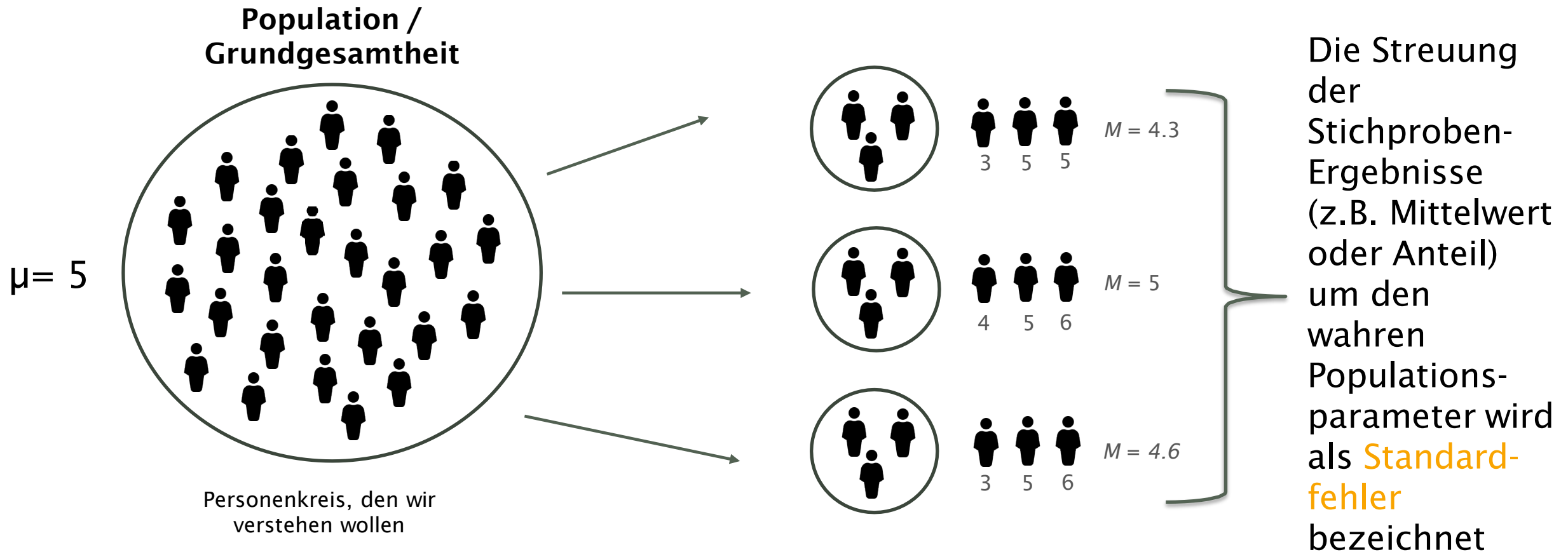
# Signifikanztests

Hypothesentests überprüfen Aussagen über Parameter (Schätzwerte) von Grundgesamtheiten. Dabei wird von den Stichprobenergebnissen auf die Population geschlossen.



# Signifikanztests

Die Ergebnisse einer Stichprobe können nicht einfach auf die Population übertragen werden, weil Stichprobenergebnisse variieren



# Signifikanztests

Ein Signifikanztest beantwortet die folgende Frage

***Wie wahrscheinlich ist es, dass das gefundene Stichprobenergebnis nur durch Zufall zustande gekommen ist und in der Population gar nicht existiert?***

# Signifikanztests

## ► Aufbau eines Signifikanztests:

Ein Signifikanztest basiert auf einer **Nullhypothese  $H_0$**  und einer **Alternativhypothese  $H_1$** :

Die **Alternativhypothese  $H_1$**  entspricht unserer Arbeitshypothese und postuliert das, was wir eigentlich zeigen wollen:

*z.B. Es gibt in der Population einen Unterschied zwischen dem Durchschnittsverdienst von Männern und Frauen  $\rightarrow H_1: \mu_{\text{Männer}} \neq \mu_{\text{Frauen}}$*

Die **Nullhypothese  $H_0$**  postuliert das Gegenteil:

*z.B. Es gibt in der Population keinen Unterschied im Durchschnittsverdienst zwischen Männern und Frauen  $\rightarrow H_0: \mu_{\text{Männer}} = \mu_{\text{Frauen}}$*   
***Das Stichprobenergebnis ist durch Zufallsvariation zustande gekommen.***



# Hypothesentests sind sogenannte Signifikanztests

- ▶ Signifikanztests weisen als Hauptergebnis einen **p-Wert** aus.

Der  $p$ -Wert gibt an, wie wahrscheinlich es ist das gefundene Stichprobenergebnis zu erhalten, **durch Zufallsvariation zustande gekommen ist.**

Ein kleiner  $p$ -Wert als Ergebnis eines Signifikanztests ist also wünschenswert, um die Nullhypothese ablehnen zu können.

Doch bis zu welchem  $p$ -Wert kann die  $H_0$ -Hypothese abgelehnt werden?

*In den empirischen Sozialwissenschaften akzeptiert man üblicherweise eine Irrtumswahrscheinlichkeit  $\alpha$  (=die Wahrscheinlichkeit, dass man die  $H_0$ -Hypothese fälschlicherweise ablehnt) von 5%. Wenn der  $p$ -Wert also  $< \alpha$  (und damit  $< 0.05$ ) ist, können wir die  $H_0$ -Hypothese ablehnen und davon ausgehen, dass unsere  $H_1$  zutrifft.*

# Agenda

- ▶ Multi-Item Skalen
- ▶ Festlegen der geeigneten statistischen Methode
  - ▶ Hypothesentests als Signifikanztests
  - ▶ **t-Test**
  - ▶ Einfaktorielle ANOVA (1 UV mit 2 oder mehr Stufen)

## T-Test

Test auf Signifikanz von Unterschieden zwischen  
2 Gruppenmittelwerten

Für unabhängige  
Stichproben

Ergeben sich z.B. aus  
Experimenten mit „**between-  
subjects-Design**“  
(Vergleich verschiedener  
Teilnehmer miteinander)

Für verbundene  
Stichproben

Ergeben sich z.B. aus  
Experimenten mit „**within-  
subjects-Design**“  
(Vergleich gleicher Teilnehmer  
miteinander)

# Vergleich von zwei Mittelwerten mit t-Test für unabhängige Stichproben

Welche Preisaktion ist effektiver?

Gruppe 1:  
Kauf zwei, bekomme eine gratis!



Messung der Kaufabsicht  
1 = auf keinen Fall .... 7 =  
sehr wahrscheinlich

**$M = 5.3$**

Wurde in einem  
Experiment im  
between-subjects-  
Design getestet

Ergibt zwei  
unabhängige Gruppen

Gruppe 2:  
Kauf drei für den Preis von 2



Messung der Kaufabsicht  
1 = auf keinen Fall .... 7 =  
sehr wahrscheinlich

**$M = 4.8$**

# Vergleich von zwei Mittelwerten – t-Test

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Gleichheit der Mittelwerte						
		F	Sig.	t	df	Sig. (2-seitig)	Mittlere Differenz	Stdfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
Kaufbereitschaft	Varianzen sind gleich	.08	.776	-2.43	75.00	.017	-.60	.25	-1.09	-.11
	Varianzen sind nicht gleich			-2.43	74.72	.018	-.60	.25	-1.09	-.11

Der Levene-Test auf Varianzgleichheit überprüft, ob die Gruppenvarianzen gleich sind (eine der Voraussetzungen für den t-Test).

Ist der p-Wert  $> .05$  kann die Nullhypothese nicht abgelehnt und es muss angenommen werden, dass die Gruppenvarianzen gleich sind. In dem Fall sind alle relevanten Werte der oberen Zeile „Varianzen sind gleich“ zu entnehmen. Ist der p-Wert  $< .05$ , können die „korrigierten“ Werte aus der unteren Zeile abgelesen werden.

Der p-Wert liegt unter dem Signifikanzniveau von  $\alpha=0.05$ .

$H_0$  (es gibt keinen Unterschied zwischen den Gruppen) wird verworfen,  $H_1$  (es gibt **in der Population** einen Unterschied zwischen den Gruppen) gilt damit als bestätigt.

# Wie man die Ergebnisse eines t-Tests berichtet

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit				T-Test für die Gleichheit der Mittelwerte				
		F	Sig.	t	df	Sig. (2-seitig)	Mittlere Differenz	Stdfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
Kaufbereitschaft	Varianzen sind gleich	.08	.776	-2.43	75.00	.017	-.60	.25	-1.09	-.11
	Varianzen sind nicht gleich			-2.43	74.72	.018	-.60	.25	-1.09	-.11

Im Durchschnitt gaben die Befragten eine höhere Kaufbereitschaft bei der Preisaktion «Kauf zwei, erhalte eine gratis» an ( $M = 5.3$ ), als bei der Preisaktion «Kaufe drei für den Preis von zwei» ( $M = 4.8$ ). Dieser Unterschied war statistisch signifikant ( $t(75) = -2.43, p = .017$ )

# Agenda

- ▶ Multi-Item Skalen
- ▶ Festlegen der geeigneten statistischen Methode
  - ▶ Hypothesentests als Signifikanztests
  - ▶ t-Test
  - ▶ Einfaktorielle ANOVA (1 UV mit 2 oder mehr Stufen)

# Vergleich von 3 Gruppenmittelwerten

$H_1$ : Mindestens zwei der drei Gruppenmittelwerte unterscheiden sich

$H_0$ : Alle drei Gruppenmittelwerte sind gleich

Gruppe 1:

Kauf zwei, bekomme eine gratis!



Messung der Kaufabsicht

1 = auf keinen Fall .... 7 = sehr  
wahrscheinlich

$M = 3.7$

Gruppe 2:

Kauf drei für den Preis von zwei!



Messung der Kaufabsicht

1 = auf keinen Fall .... 7 = sehr  
wahrscheinlich

$M = 2.63$

Gruppe 3:

Alle guten Dinge sind drei!



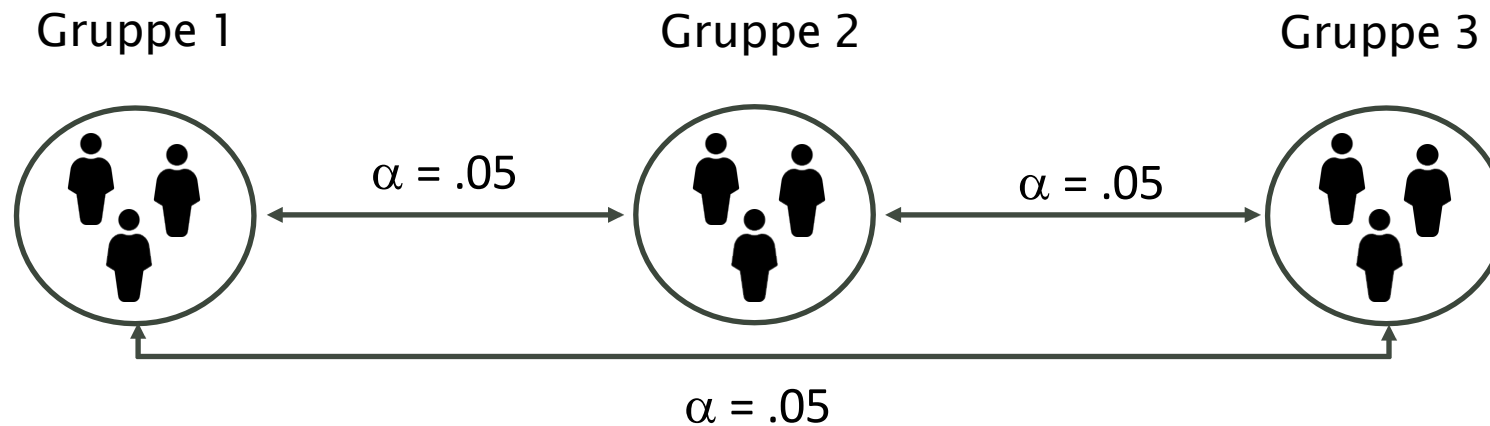
Messung der Kaufabsicht

1 = auf keinen Fall .... 7 = sehr  
wahrscheinlich

$M = 3.59$

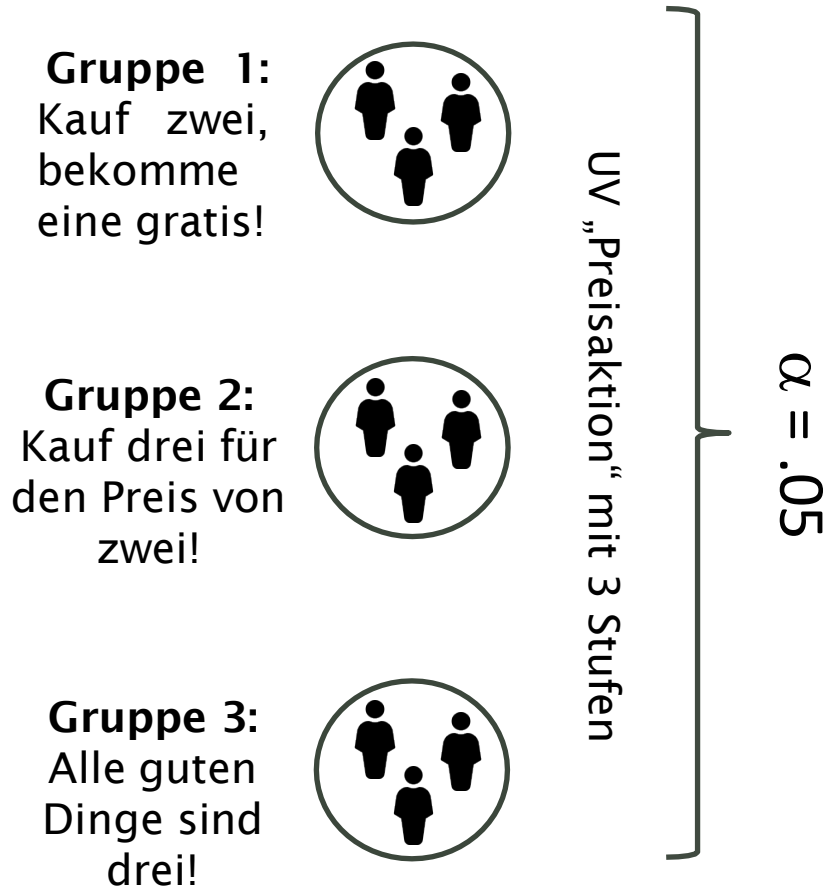


# Warum können wir den t-Test in diesem Fall nicht anwenden?



- Bei der Durchführung von mehreren t-Tests kumuliert sich die Irrtumswahrscheinlichkeit  $\alpha$
- Die Wahrscheinlichkeit, dass unser Befund tatsächlich statistisch signifikant ist, beträgt  $1 - \alpha = .95$
- Bei drei t-Tests beträgt diese Wahrscheinlichkeit  $(1 - \alpha) * (1 - \alpha) * (1 - \alpha) = .95 * .95 * .95 = .86$
- Somit beträgt die kumulierte Irrtumswahrscheinlichkeit bei 3 t-Tests  $1 - .86 = .14$

# Bei Vergleichen von zwei oder mehr Gruppen aus einer UV → Einfaktorielle ANOVA



- Die ANOVA testet mit der F-Statistik alle Gruppenmittelwerte simultan und es gibt nur eine Irrtumswahrscheinlichkeit  $\alpha$

- ANOVA = Analysis of Variance**

**Varianz in der AV (Y)** = Quadratsumme der Abweichungen aller Beobachtungswerte  $y$  vom Mittelwert  $\bar{y}$

ANOVA zerlegt die **Gesamtvarianz der AV** in zwei Teile: **Varianz zwischen den Gruppen** und **Varianz innerhalb von Gruppen** und setzt diese für den Sig.-Test ins Verhältnis

$$F = \frac{\text{Varianz zwischen Gruppen}}{\text{Varianz innerhalb Gruppen}}$$

Steht für  
Unterschiede  
zwischen den  
Gruppen

- Das Ergebnis der ANOVA (F-Wert und p-Wert) sagt uns zunächst nur, **ob** ein signifikanter Unterschied vorliegt, aber **nicht welche Gruppen genau** sich unterscheiden
- Um zu bestimmen, welche Gruppenmittelwerte sich unterscheiden, muss ein **Post-hoc-Test** eingesetzt werden: Scheffé-, Tukey-, Newman-Keuls-, oder Duncan-Test.

# Wie man die Ergebnisse einer einfaktoriellen ANOVA berichtet

Deskriptive Statistiken

Gruppe		N	Mittelwert	Std. Abweichung	Standardfehler	95% Konfidenzintervall für Mittelwert		Minimum	Maximum
						Untere Grenze	Obere Grenze		
Kaufbereitschaft	Preisaktion 1	44	3.70	1.84	.28	3.15	4.26	1.00	7.00
	Preisaktion 2	43	2.63	1.38	.21	2.20	3.05	1.00	7.00
	Preisaktion 3	44	3.59	1.74	.26	3.06	4.12	1.00	6.00
	Gesamt	131	3.31	1.72	.15	3.02	3.61	1.00	7.00

ANOVA

		Quadratsumme	df	Mittel der Quadrate	F	Sig.
Kaufbereitschaft	Zwischen Gruppen	30.33	2	15.16	5.45	.005
	Innerhalb Gruppen	355.84	128	2.78		
	Gesamt	386.17	130			

Mehrfachvergleiche (Kaufbereitschaft)

						95% Konfidenzintervall	
	(J) Familie	(J) Familie	Differenz der Mittelwerte (I - J)	Standardfehler	Sig.	Untere Grenze	Obere Grenze
LSD	Preisaktion 1	Preisaktion 2	1.08	.36	.003	.37	1.78
		Preisaktion 3	.11	.36	.750	-.59	.82
	Preisaktion 2	Preisaktion 1	-1.08	.36	.003	-1.78	-.37
		Preisaktion 3	-.96	.36	.008	-1.67	-.26
	Preisaktion 3	Preisaktion 1	-.11	.36	.750	-.82	.59
		Preisaktion 2	.96	.36	.008	.26	1.67

$$F = \frac{\text{Varianz zwischen Gruppen}}{\text{Varianz innerhalb Gruppen}}$$

$$p = 0.005$$

$$df_{zw} = k - 1$$

$$df_{inn} = N - k$$

df=Freiheitsgrade, d.h. Anzahl der Werte im statistischen Ausdruck, die frei variieren können

Das Ergebnis einer einfaktoriellen Varianzanalyse zeigte, dass die Preisaktion einen signifikanten Effekt auf die angegebene Kaufbereitschaft hatte  $F(2,128) = 5.45, p = .005$ .

Post-hoc-Analysen unter der Verwendung des Sheffé Post-hoc-Kriteriums für Signifikanz zeigten, dass die Kaufbereitschaft bei Preisaktion 2 ( $M = 2.63$ ) signifikant geringer war als bei Preisaktion 1

Berner Fachhochschule ( $M = 3.70, p = .003$ ) und als bei Preisaktion 3 ( $M = 3.59, p = .008$ ).